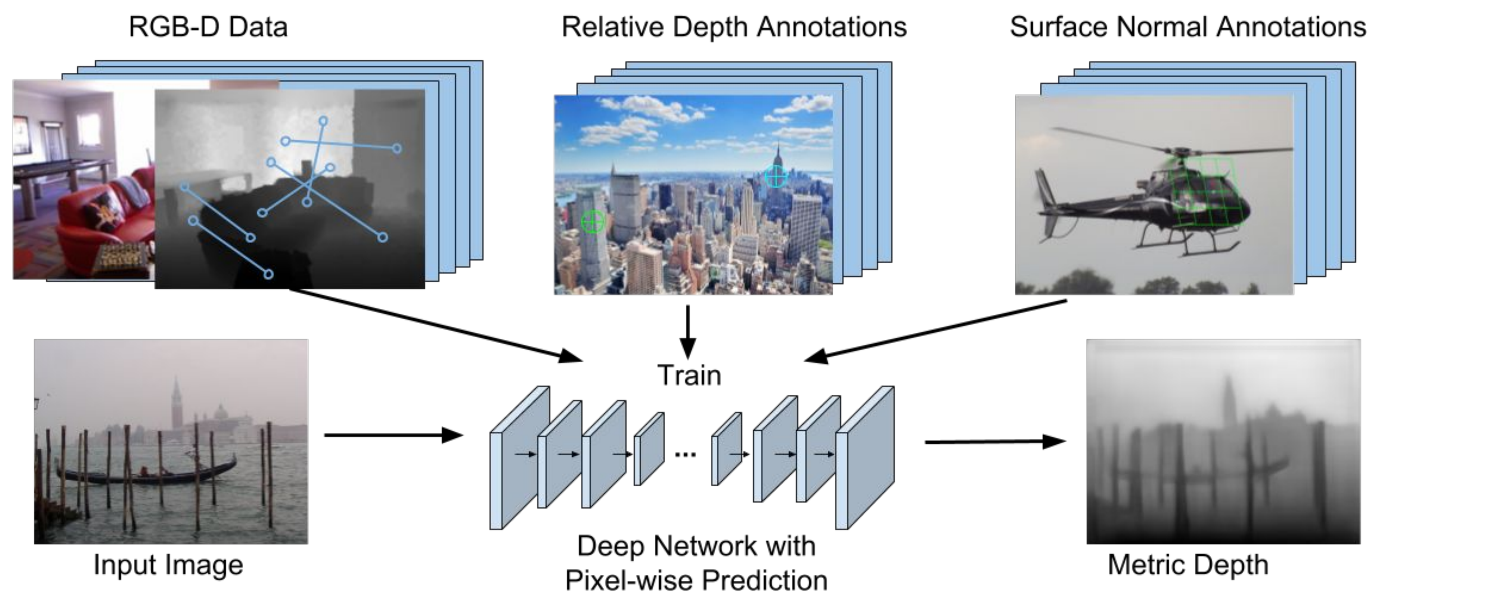
Surface Normals in the Wild

Weifeng Chen[1], Donglai Xiang[2], Jia Deng[1]

[1]University of Michigan, Ann Arbor, USA

[2]Tsinghua University, Beijing, China

Introduction



Contribution

- A new dataset of surface normals for images in the wild.
- Two distinct approaches of using surface normals + relative depth to train a depth-prediction network.

Background

Relative depth: Which is closer? point A or point B?
We can train depth-prediction networks with relative depth.

Why do we need surface normals?

Relative depths introduce ambiguities:

- Not affected by Bending/wiggling/tilting (**Figure 1**).
- Can't train continuity, surface orientation, and curvature.

Surface normal encodes orientation of surface and the derivative of depth --> eliminates ambiguities

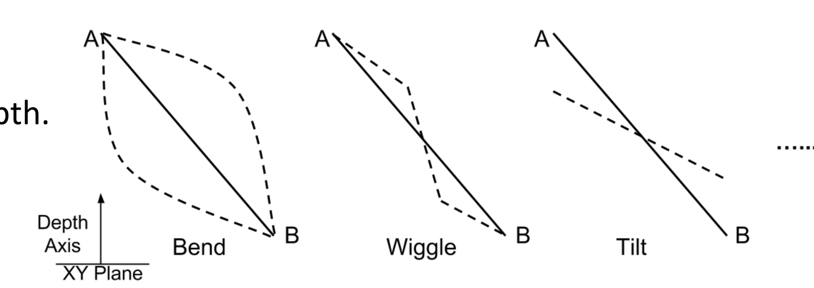


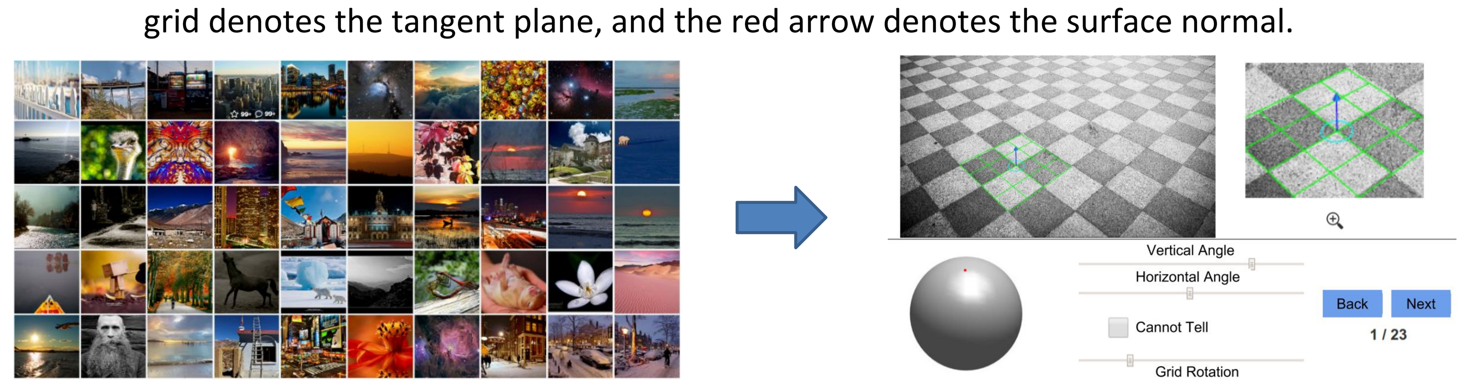
Figure 1. Bending, wiggling, or tilting does not change relative depth of point A and B.

The Surface Normals in the Wild (SNOW) Dataset

About the Dataset

- An image dataset that consists of 60,061 diverse images
- Each image comes with one randomly sampled point and its ***surface normal annotation***.

Figure 2. Examples of surface normal annotations from the SNOW dataset. The green grid denotes the tangent plane, and the red arrow denotes the surface normal.



Random Images from Flickr **Annotation UI**

Figure 3. The data collection pipeline.

Quality of human annotated surface normals

We test on 113 samples from the NYU Depth dataset, and evaluate these metrics:

- **Human-Human Disagreement (HHD)**: difference between a human annotation and the mean of multiple human annotations.
- **Human-Kinect Disagreement (HKD)**: the average angular difference between a human annotation and the Kinect ground truth.

	HHD	HKD
w/- Kinect error	7.4°	32.8°
w/o Kinect error	7.17°	15.64°

Table 1. Annotation errors on NYU Depth Dataset.

Source of error

- Holes in the Kinect raw depth map. (**Figure 4**)
- Imperfect normal computed from Kinect depth.

Result (**Table 1**)

Human annotations of surface normals are of high quality.

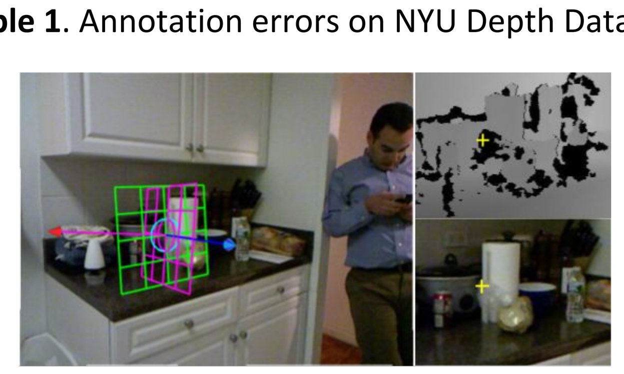


Figure 4. One example of Kinect error.

Code & Data are Available! **http://www-personal.umich.edu/~wfchen/surface-normals-in-the-wild/**

Learning with Surface Normals

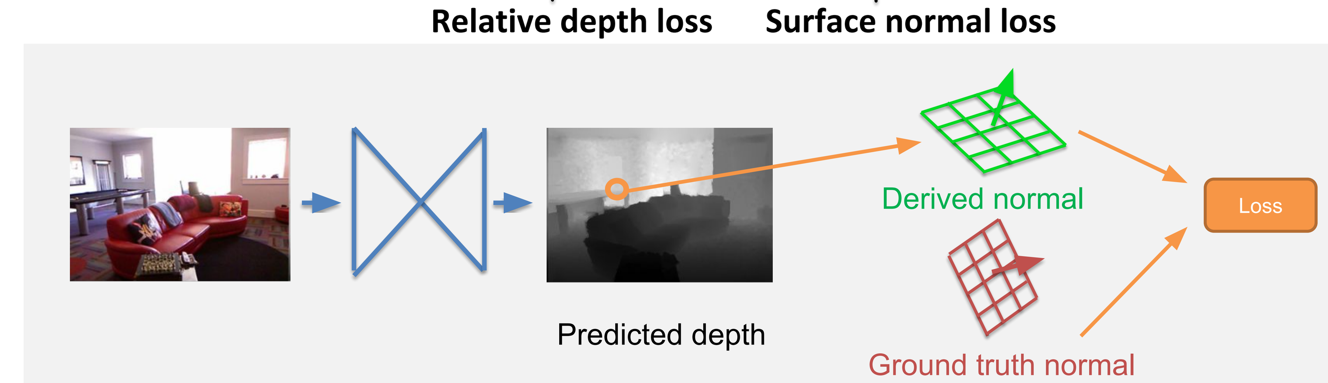
A training image *I* and its *K* queries $R = \{ (i_k, j_k, r_k)\}, k = 1, \ldots, K$, and *L* surface normal annotations $S = \{ p_l, n_l\}, l = 1, \ldots, L$

- i_k, j_k: the location of the 2 points in the k-th query,
- $r_k \in \{+1, -1, 0\}$: ground-truth depth relation between i_k and j_k -- closer (+1), further (-1), equal (0).
- z_{ik}, z_{jk} : the depths at location i_k and j_k.
- p_l, n_l: the location of the l-th annotation and the groundtruth surface normal at p_l.

Overall Loss function

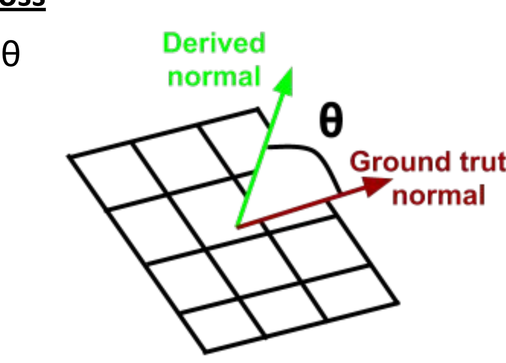
Encourage the predicted depth to be consistent with both the ground truth relative depth and the ground truth surface normals:

$$L(R,S,z) = \underbrace{\frac{1}{K}\sum_{k=1}^{K} \psi(i_k, j_k, r_k, z)}_{\text{Relative depth loss}} + \lambda \underbrace{\frac{1}{L}\sum_{l=1}^{L} \phi\,(p_l, n_l, z)}_{\text{Surface normal loss}}$$



Angle-based surface normal loss

The difference in orientation: θ



Depth-based surface normal loss

Idea: compute the "should-be" depth value of a neighbor using the ground truth normal, and penalize its difference with the predicted depth.

$$\phi(p_l, n_l, z) = \sum_{i \in \{T,B,L,R\}} \left(\hat{z}_{p_l^i} - z_{p_l}\right)^2 / \left(\hat{z}_{p_l^i} + z_{p_l}\right)^2$$

- z_{p_l}: the predicted depth at location p_l
- $\hat{z}_{p_l^i}$: be "should-be" depth at location p_l generated by the predicted depth on Top/Bottom/Left/Right of p_l.

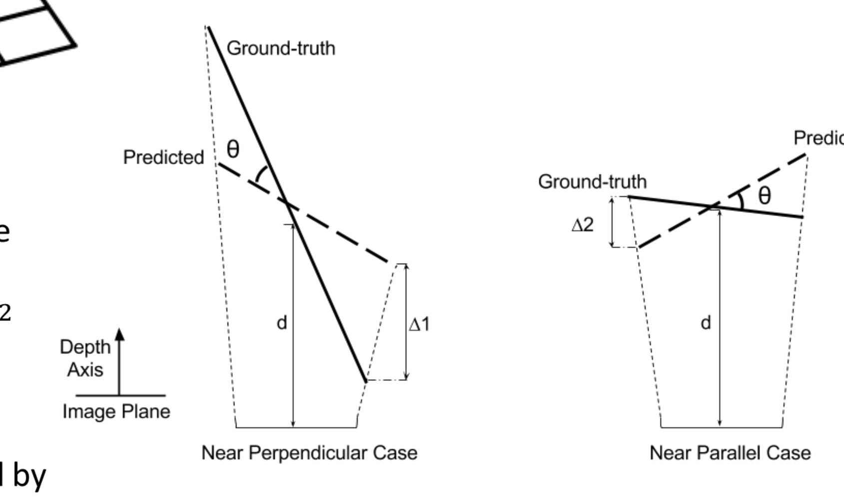


Figure 5. Two 3D planes (solid line). The predicted planes (dotted lines) both deviate by θ from the ground-truth, but incur drastically different metric depth errors Δ1 and Δ2.

Experiments

A. Experiments on NYU Depth & KITTI

Experimental Setup:

We compare these 3 models on the NYU and KITTI:

- ***d:*** model trained with relative depth
- ***d_n_al:*** relative depth + surface normal using angle-based loss
- ***d_n_dl:*** relative depth + surface normal using depth-based loss

Normal Error Evaluation Metric:

- Mean & median of angular difference with the ground-truth
- Percentages of predicted samples who are within *t* degrees of the ground-truth.

Surface normals are generated **from the predicted depth.**

Depth Error Evaluation Metric:

- **WKDR**: the overall disagreement rate between the predicted ordinal relations and ground-truth ordinal relations.
- **WKDR$^{=}$** : WKDR on pairs whose ground-truth relations are = .
- **WKDR$^{\neq}$** : WKDR on pairs whose ground-truth relations are < or >.
- **RMSE, log RMSE, etc:** Normalized to have the same mean and standard deviation as those of the mean depth map of the training set.
- **LS_RMSE**: least squared differences under a global scaling and translation of the depth values:

$$LS\_RMSE(z, z^*) = \min_{a,b} \sum_i (az_i + b - z_i^*)^2$$

Results (Table 2 , 3 , 4 ,5 & 6)

- **Depth-based loss:** Significant improvement in metric depth. No significant improvement in ordinal depth. Improvement in surface normal estimation.
- **Angle-based normal loss:** Not so significant improvement in metric depth. Better ordinal depth. Outperforms all other methods on surface normal estimation.
- The two losses have a different set of tradeoffs and are appropriate in different applications.

Method	RMSE	RMSE (log)	RMSE (s.inv)	absrel	sqrrel	LS RMSE
d	1.08	0.37	0.23	0.34	0.41	0.52
d_n_al	1.09	0.38	0.24	0.34	0.42	0.55
d_n_dl	**1.08**	**0.37**	**0.23**	**0.34**	**0.41**	**0.50**
Chen_Full[1]	1.11	0.38	0.24	0.34	0.42	0.58
Eigen(V)*[2]	0.64	0.21	0.17	0.16	0.12	0.47

Table 2. Metric depth error on the NYU Depth dataset. Eigen(V)* is trained on full metric depth.

Model	Angle Distance		% Within t°		
	Mean	Median	11.25	22.5	30
d	29.45	22.71	22.31	50.71	63.65
d_n_al	**25.92**	**20.09**	**26.28**	**56.45**	**69.26**
d_n_dl	30.85	24.51	24.51	46.93	60.31
Chen_Full[1]	30.35	24.37	18.64	46.80	61.42
Eigen(V)[2]	35.97	28.34	17.67	41.12	53.49

Table 3. Surface normal error evaluated on the NYU Depth dataset.

Method	RMSE	RMSE (log)	RMSE (s.inv)	absrel	sqrrel	LS RMSE
d	6.86	2.06	1.92	0.38	2.77	5.66
d_n_al	6.75	1.56	1.45	0.34	2.45	5.57
d_n_dl	6.17	0.83	0.76	0.28	1.88	4.84
Godard[4]	**5.21**	**0.22**	**0.20**	**0.11**	**0.89**	**4.73**

Table 4. Metric depth error on the KITTI dataset.

Method	WKDR	WKDR=	WKDR≠
d	29.2%	32.5%	28.0%
d_n_al	**27.6%**	**31.5%**	**26.6%**
d_n_dl	30.9%	31.7%	31.4%
Chen_Full[1]	28.3%	30.6%	28.6%
Eigen(V)*[2]	34.0%	43.3%	29.6%

Table 5. The ordinal depth error on the NYU Depth dataset.

Method	WKDR	WKDR=	WKDR≠
d	26.46%	24.01%	27.08%
d_n_al	**22.35%**	**20.61%**	**22.93%**
d_n_dl	26.50%	22.58%	27.50%
Godard[4]	25.84%	26.17%	26.21%

Table 6. The ordinal depth error on the KITTI Depth dataset.

B. Experiments on Surface Normals in the Wild (SNOW)

	Model	Angle Distance		% Within t°		
		Mean	Median	11.25	22.5	30
Normals From Predicted Depth	d_n_al	32.53	27.44	15.40	40.52	54.12
	d_n_al_SNOW	**25.75**	**21.26**	**21.66**	**52.98**	**67.88**
	Chen_Full	35.16	30.26	13.70	36.56	49.56
	Eigen(V)[2]	48.71	46.15	6.35	18.91	28.45
Directly Predicted Normals	Ours_NYU§	31.96	26.03	18.16	43.72	56.03
	Ours_NYU_SNOW§	**23.33**	**17.99**	**30.42**	**60.54**	**72.74**
	Eigen(V)[2] §	28.71	23.16	20.98	48.78	61.84
	Bansal[3]§	27.85	22.25	23.41	50.54	64.09

Table 7. Surface normal error evaluated on SNOW. Models with a **§** suffix directly predict surface normals.

- ***d_n_al_F_SNOW:*** d_n_al_F fine-tuned on SNOW. Normal from depth.
- ***Ours_NYU:*** Network trained on NYU directly predicts surface normal.
- ***Eigen/Chen_Full:*** Baselines trained on NYU. Normal from depth.
- ***Ours_NYU_SNOW:*** Ours_NYU fine-tuned on SNOW. Normal from depth.
- ***Bansal***: Baseline network trained on NYU directly predicts surface normal.

Experimental Setup:

Train/test split: 49,805 training, 10,256 test.

Results (Table 7)

- ***d_n_al_F_SNOW*** and ***d_n_al_F_SNOW§*** achieve the best result.

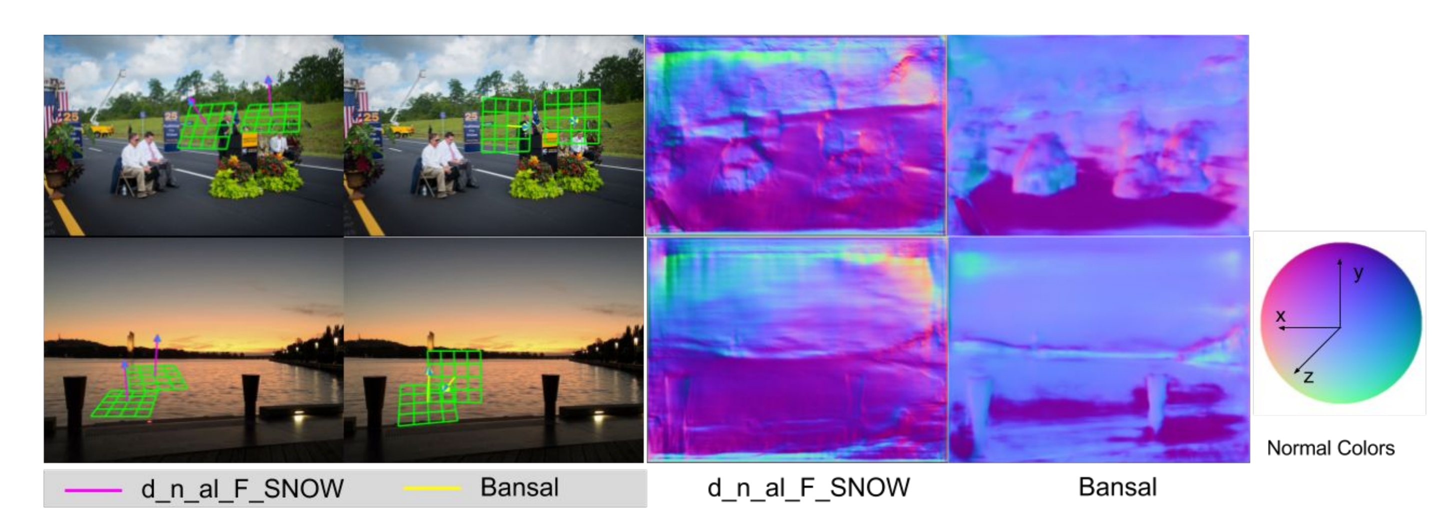


Figure 6. Qualitative results on SNOW produced by our model and Bansal

References

[1] Chen, Weifeng, et al. "Single-image depth perception in the wild." In NIPS. 2016.

[2] Eigen et al. "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture." In ICCV. 2015.

[3] Bansal et al.. "Marr revisited: 2d-3d alignment via surface normal prediction." In CVPR. 2016.

[4] Godard et al. "Unsupervised monocular depth estimation with left-right consistency." arXiv preprint arXiv:1609.03677 (2016).